

La codifica XML/TEI



DTD/Schema TEI P5

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html>

TEI Consortium, eds.
Guidelines for Electronic Text Encoding and Interchange.

<http://www.tei-c.org/P5/>

**Linee guida per la codifica e l'interscambio
del testo elettronico**

Il ruolo dell'XML

<http://www.w3.org/XML>

- Metalinguaggio = norme **SINTATTICHE**
 - [Extensible Markup Language \(XML\) 1.0 \(Fifth Edition\)](#)
W3C Recommendation 26 November 2008
- e cioè assenza di un **vocabolario**
- Concetto di *Tipo di Documento* (la **classe**).
- Non solo concetto tecnico: la **DTD**
(<http://www.w3schools.com/dtd/default.asp>)
 - ma ora
- Gli *Schemi* (<http://www.w3schools.com/schema>)
- Importanza del concetto di **MODELLO**
 - Fra schema e processo di astrazione delle proprietà pertinenti
- La **struttura gerarchica ad albero**

2. VESPASIANO A FILIPPO PODOCATARO

Firenze, 4 dicembre 1448

Firenze, Bibl. Nazionale, Magl. VIII, 1390, f. 102

Vespasianus Philippo Podaghatero Cyprio plurimam salutem.

Vehementer me oblectant [...]

Vale. Ex Florentia, die 4 decembris 1448.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<lettera type="inviata" n="2">
  <metadati>
    <identificativo>2</identificativo>
    <mittente>VESPASIANO</mittente>
    <destinatario>
      <persona riferimento="#FiPo" chiave="Podocataro, Filippo">
        <link rif="&Podocataro;">FILIPPO PODOCATARO</link>
      </persona>
    </destinatario>
    <datazione>
      <data_topica>Firenze</data_topica>
      <data_cronica>4 dicembre 1448</data_cronica>
    </datazione>
    <descrizione_catalografica>
      <luogo>Firenze</luogo>,
      <istituzione>Bibl. Nazionale</istituzione>
      <segnatura>Magl. VIII, 1390 <carte> f. 102</carte></segnatura>
    </descrizione_catalografica>
  </metadati>
  <testo_lettera>
    <incipit>Vespasianus Philippo Podaghatero Cyprio plurimam salutem. </incipit>
    <corpo>Vehementer me oblectant [...] </corpo>
    <explicit type="salutatio">&V;</explicit>.
    <explicit type="datatio">Ex Florentia, die 4 decembris 1448.</explicit>
  </testo_lettera>
</lettera>
```

Un documento XML "ben formato"

Nessuno schema di riferimento, ma un modello interpretativo, esito di un processo di astrazione delle proprietà pertinenti

La modellazione si riflette nella struttura gerarchica ad albero della rappresentazione del documento

L'importanza del modello per i testi umanistici

La definizione del modello deriva da due precise circostanze:

- specificità della risorsa, nei termini di **tipologia documentaria**,
- **caratteristiche dell'analisi** e del successivo trattamento automatico del testo o del *corpus*.

Ogni modello del testo è dunque il risultato dell'assunzione di un **punto di vista** sulla fonte.

I macrolivelli interpretativi

Potremmo ipotizzare tre livelli, che corrispondono a tre macro aree di analisi e quindi a tre modelli o a tre punti di vista:

- STRUTTURA
- RAPPRESENTAZIONE/FORMATO
- FENOMENI INTERLINEARI/CONTENUTO

Non è obbligatorio scegliere uno dei tre modelli; tutti e tre possono essere rappresentati in fase di *markup* o meglio il modello elaborato può contemplare elementi che provengono dai tre diversi livelli della modellizzazione.

Struttura

- Per codifica della **struttura** si intende l'individuazione delle partizioni principali e secondarie del testo.
- Ci si riferisce, ad esempio, alle suddivisioni formali in capitoli, sezioni, paragrafi per un testo in prosa; canti, stanze, versi per un componimento poetico; atti e scene per un testo teatrale, ecc.

Rappresentazione/formato

- Si intende l'aspetto fisico del documento, quindi la riproduzione in formato digitale delle caratteristiche della versione cartacea.
 - grassetto, corsivo e sottolineato,
 - caratteri speciali che richiedono una particolare resa per essere rappresentati (tutto ciò che esula quindi dal set ASCII base o dall'Unicode UTF 8),
 - posizionamento di note e annotazioni, organizzazione di eventuali glosse marginali,
 - indentazioni,
 - a capo,
 - caratteristiche materiali (come aggiunte e cancellazioni),
 - alternanza maiuscole/minuscole, interpunzione.
- Gli **aspetto di formato**, in particolare, sono caratteristiche che possono avere un ruolo importante nell'**interpretazione del significato**: un corsivo può qualificare una parola in lingua straniera, un titolo, un' enfasi nel contenuto e non essere un mero accorgimento tipografico.

Fenomeni interlineari/Contenuto

- In questa fase il *markup* richiederà una scelta specifica dei fenomeni in relazione alle **esigenze analitiche del lavoro di resa elettronica**.
- Qualche esempio:
 - nomi di persona e riferimenti di luogo,
 - numeri e date,
 - citazioni,
 - figure retoriche,
 - lessico tecnico/specifico,
 - parole in lingua straniera,
 - analisi stilistica,
 - codifica delle varianti all'interno di una tradizione testuale,
 - annotazioni morfo-sintattiche e semantiche di *corpora*.

Il ruolo della TEI

Per agevolare la **portabilità** dei prodotti digitali e per consentire dunque la **condivisione dei formati di scambio** è stato realizzato un progetto denominato TEI (*Text Encoding Initiative*).

La TEI è una DTD (ora anche XML schema) o modello di codifica che vuole contemplare tutta la serie dei fenomeni di interesse 'umanistico' e trovare, per ciascuno di essi, un **vocabolario unico** al fine di arrivare ad una formalizzazione utile a normalizzare i **criteri, le modalità e il lessico** del *markup*, di fronte alla polisemia del linguaggio naturale.

Potremmo definire la TEI come una **grammatica** avente il compito di assegnare un **vocabolario controllato** all'XML.

Termini e relazioni di dipendenza gerarchica.

Origini della TEI

- A partire dal 1987 le tre maggiori associazioni mondiali di studiosi di scienze umane attraverso metodologie informatiche, la *Association for Computers and the Humanities* (ACH) la *Association for Computational Linguistics* (ACL) e la *Association for Literary and Linguistic Computing* (ALLC) hanno avviato un progetto internazionale per sviluppare un **modello di codifica normalizzato**.
- Questo progetto è stato denominato *Text Encoding Initiative* (TEI), sito ufficiale <http://www.tei-c.org/>.
- Ad esso hanno attivamente collaborato un gruppo di studiosi provenienti da differenti paesi

Guidelines TEI

- Il lavoro delle commissioni, iniziato nel 1989, ha condotto alla realizzazione di una vasta e complessa DTD (*Document Type Definition*), le cui caratteristiche sono state per la prima volta presentate nella pubblicazione, uscita nel 1994, col titolo *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*.
- Nel 2002 è stata rilasciata la TEI P4, per venire incontro all'esigenza di prevedere, anche nelle *Guidelines*, l'adeguamento della rinnovata DTD alla sintassi XML.
- Nel 2007 è stata rilasciata la TEI P5, che ha comportato una sostanziale revisione delle *TEI Guidelines* (compresa la realizzazione di una versione della TEI basata su XML Schema che si affianca alla DTD).

Cfr. <http://www.tei-c.org/Guidelines>

La funzione delle *Guidelines*

- Sono indirizzate a tutti coloro che intendono scambiare **informazioni archiviate in formato elettronico**.
- Sottolineano l'importanza dello scambio di informazioni testuali, ma trattano anche di **altre forme di informazioni** (quali immagini e suoni).
- Sono applicabili indifferentemente sia per la **creazione di nuove risorse** che per lo **scambio** di quelle già esistenti.
- Forniscono un mezzo per **rendere esplicite certe caratteristiche di un testo** in modo tale da facilitarne il trattamento mediante programmi basati su diverse piattaforme.
- Qualsiasi rappresentazione di un testo su un computer usa una qualche forma di codifica; la TEI è stata creata sia per **ovviare alla eccessiva varietà di schemi di codifica** tra loro incompatibili (interscambio), sia per il crescente numero di **applicazioni utilizzate per i testi in formato elettronico**.

Textual Phenomena

The TEI Guidelines define a set of rules to mark up the phenomena in a wide range of texts and textual objects in a descriptive fashion.

Generally speaking, there are four classes of textual phenomena that can be described:

- Structural
- Renditional
- Logical & Semantic
- Analytic

Structural and renditional features are best understood because they concern a natural kind of textual, though culturally defined, organisation.

Books mainly consist of chapters, sections, and paragraphs; poetry is mostly organised in poems, stanzas, and lines; whereas scenes, acts, and parts of speech are **structural features** of performance texts.

In these texts, linguistic units are **highlighted** by the use of distinct fonts, colours, alignments, italics, underlinings, font weight etc. These textual codes signal underlying **logical and semantic features** and functions such as names of organisations, titles of books, distinctive languages, emphatic language use, etc.

However, semantic and logical features don't need to be highlighted by means of typographic codes and can occur in texts unsuspectingly. It needs a thorough understanding of the text and the language to identify them.

Semantic and syntactic interpretations added to a text or part of a text that together constitute a new text, we call analytical features. Examples are linguistic (wordclass, morpheme,...) and narrative (theme, motive,...) categorisations.

Tratto da: *TEI by example*, <http://tbe.kantl.be/TBE/modules/TBED01v00.htm#textualphenomena>

TEI *Lite*

È stato elaborato anche un sottoinsieme della TEI, denominato *TEI Lite*, che dovrebbe consentire la diffusione del progetto di codifica, senza impegnare l'utenza ad apprendere lo schema nella sua interezza.

Si tratta di una versione semplificata dell'intero schema di codifica che permette di facilitare la realizzazione di testi in formato elettronico *tei-compliant*.

Elenco dei marcatori e degli attributi [TEI-Lite](#)
(140 circa)

Struttura TEI: testo e metadati

Ogni testo codificato conformemente alle specifiche della TEI è costituito da due parti:

un *TEI header* ("testata"), contenente le informazioni editoriali concernenti, a diverso livello, il documento elettronico;

un *TEI text* contenente la trascrizione codificata del testo in versione integrale.

[Template TEI](#)

TEI *header*

Diremo che nella “testata” della TEI sono raccolte tutte le informazioni utili per:

- descrizione bibliografica del testo elettronico e del suo esemplare originale di riferimento,
- modalità e caratteristiche della codifica,
- note non bibliografiche e revisioni.

Queste notizie possono definirsi come i metadati della risorsa, cioè le meta-informazioni utili a descrivere il testo elettronico prodotto e il documento materiale utilizzato come esemplare per la trascrizione digitale.

Un esempio da [Biblioteca Italiana](#)

Le opere della tradizione culturale italiana in formato digitale testuale

Il Frontespizio elettronico

Ogni testo TEI ha una testata che offre informazioni analoghe a quelle fornite dal frontespizio di un testo a stampa. La testata è introdotta dall'elemento *<teiHeader>* ed è composta da quattro parti principali:

1. *<fileDesc>* contiene una descrizione bibliografica completa di un file digitale.
2. *<encodingDesc>* documenta le relazioni tra un testo elettronico e la fonte, o le fonti, da cui è stato tratto.
3. *<profileDesc>* contiene una descrizione dettagliata degli aspetti non bibliografici di un testo, specificamente le lingue e i dialetti usati, le circostanze in cui è stato prodotto, i partecipanti e il loro ambiente.
4. *<revisionDesc>* riassume la storia delle revisioni di un documento elettronico.

Es. offline: [Dante, De vulgari eloquentia](#)

TEI *text*

L'elemento *text* si divide, a sua volta, in quattro elementi:

<front> (opzionale),

<body> (obbligatorio),

<group> (obbligatorio)

<back> (opzionale)

<Front> e <back> sono necessari qualora il documento riporti delle informazioni, a livello paratestuale, all'inizio o alla fine del testo vero e proprio (per esempio all'inizio si possono trovare un indice, un'intestazione, una prefazione, una dedica, ecc.; alla fine un glossario, un indice, un'appendice, una postfazione, ecc.), delle quali si necessita operare una codifica.

<body> è l'elemento che introduce il corpo del testo.

<group> è da impiegarsi nel caso di una serie di testi facenti parte della medesima pubblicazione, ognuno dotato della propria individualità (ognuno con un proprio front, body e back).

Struttura generale TEI

Un testo unitario sarà codificato usando una struttura generale come questa:

```
<TEI.2>
  <teiHeader> [testata TEI] </teiHeader>
  <text>
    <front> [avantesto] </front>
    <body> [testo] </body>
    <back> [annessi] </back>
  </text>
<TEI.2>
```

Le suddivisioni interne del corpo del documento

Structural

All'interno dell'elemento `<body>` ci saranno ovviamente una serie di suddivisioni ulteriori, necessarie a definire la struttura interna del testo, in modo direttamente proporzionale alla tipologia (prosa, versi, testo teatrale, dizionario, parlato, ecc.).

Tali suddivisioni evidenzieranno la scansione logica del testo e la sua articolazione in sezioni di vario livello (`<div>`, `<div0>`, `<div1>`, `<div2>`), in paragrafi (`<p>`) se necessario in linee (`<l>`), titoli (`<head>`), ecc.

Saranno necessari marcatori differenti a seconda del **tipo di testo** codificato e a seconda dei **fenomeni** che si intendono marcare.

Allora alla codifica della struttura si aggiungerà quella di alcuni elementi ritenuti necessari alla descrizione della fonte o agli scopi della ricerca: nomi di persona, discorsi diretti, parole in lingua straniera, note, date, citazioni bibliografiche, liste, etc.

Elementi per la segmentazione del testo (<body>)

Il corpo di un testo in prosa può essere costituito solamente da una serie di paragrafi, oppure tali paragrafi possono essere raggruppati insieme in capitoli, sezioni, sottosezioni, etc.

Nel primo caso, ogni paragrafo è marcato usando il marcatore <*p*>.

Nel secondo caso, l'elemento <*body*> può essere diviso sia in una serie di elementi <*div1*>, sia in una serie di elementi <*div*>, i quali possono essere ulteriormente suddivisi.

Marcatore di segmentazione

`<p>` marca i paragrafi in prosa.

`<div>` contiene una sezione dell'avantesto, del corpo o degli annessi di un testo.

`<div1>` contiene una sezione di primo livello dell'avantesto, del corpo ed degli annessi di un testo (la più ampia, se `<div0>` non è usato, la seconda in ordine di grandezza altrimenti)

Qualora siano necessarie suddivisioni strutturali più piccole di una `<div1>`, quest'ultima può essere divisa in elementi `<div2>`, una `<div2>` in più piccoli elementi `<div3>`, etc., fino al livello `<div7>`. Se sono presenti più di 7 livelli della divisione strutturale, è necessario o modificare la DTD TEI per accettare `<div8>`, etc, oppure usare l'elemento non numerato `<div>`: un elemento `<div>` può essere suddiviso in elementi più piccoli `<div>`, senza limiti alla profondità di annidamento.

Attributi degli elementi di segmentazione (1)

@type

indica il nome convenzionale per questa categoria di divisione del testo. Il suo valore sarà generalmente *LIBRO*, *CAPITOLO*, *POEMA*, etc.

Altri possibili valori includono *GRUPPO* per gruppi di poemi, etc, trattati come un'unità singola, *SONETTO*, *DISCORSO* e *CANZONE*.

Si noti che qualsiasi valore venga fornito per l'attributo *type* del primo *<div>*, *<div1>*, *<div2>*, etc., in un dato testo, si assume che sia valido per tutte le successive *<div>*, *<div1>*, all'interno dello stesso *<body>*.

Questo implica che un valore deve essere fornito per il primo elemento di divisione di ogni tipo, o ogni qualvolta il valore cambi.

Attributi degli elementi di segmentazione (2)

@xml:id

specifica un identificatore unico per la divisione, che può essere usato per i riferimenti incrociati o per altri collegamenti, come una nota o un commento.

Spesso è utile dare un attributo @xml:id a ogni unità strutturale importante di un testo, e far derivare i valori in un modo sistematico, ad esempio aggiungendo un numero di sezione a un breve codice per il titolo dell'opera.

Attributi degli elementi di segmentazione (3)

@n

specifica un breve nome mnemonico o un numero per la divisione, che può essere usato per identificarla, al posto o assieme a *@xml:id* (spesso preferito per i riferimenti incrociati).

Se già esiste una forma convenzionale di riferimento o un'abbreviazione per le parti di un lavoro (come il libro/capitolo/verso delle citazioni bibliche), l'attributo *@n* è il luogo deputato a registrarlo.

Caratteristiche degli attributi

- Gli attributi *@xml:id* e *@n* sono tanto utili da essere ammessi in tutti gli elementi, in ogni DTD conforme alla TEI: sono *attributi globali*.
- Il valore di ogni attributo *@xml:id* deve essere unico all'interno di un documento.
- Un semplice modo per assicurarsi che questo avvenga è fare in modo che esso rifletta la struttura gerarchica del documento.

Esempio di codifica

(elemento <div> e attributi)

```
<div1 xml:id='RN1' n='I' type='libro'>
  <div2 xml:id='RN101' n='I.1' type='capitolo'> ... </div2>
  <div2 xml:id='RN102' n='I.2' type='capitolo'> ... </div2>
  ...
  <div2 xml:id='RN110' n='I.10' type='capitolo'>
    <div3 xml:id='RN1101' n='I.10.1' type='sezione'>... </div3>
    <div3 xml:id='RN1102' n='I.10.2' type='sezione'>... </div3>
  </div2>
  ...
</div1>
<div1 xml:id='RN2' n='II' type='libro'>
  ...
</div1>
...
```

(da *La Ricchezza delle Nazioni* di Smith)

Titoli e chiusure

Ogni `<div>`, `<div1>`, `<div2>`, etc., può avere un titolo o un'intestazione al suo inizio, e (meno comunemente) una chiusura quale 'Fine del capitolo 1'.

I seguenti elementi possono essere usati per trascriverli:

- ▣ `<head>` contiene ogni tipo di intitolazione, per esempio, il titolo di una sezione, oppure l'intestazione di una lista o di un glossario.
- ▣ `<trailer>` contiene una formula di chiusura o un elemento a piè di pagina che compare alla fine di una divisione di un testo.

Esempio:

```
<div1 xml:id='UNC1' n='Libro primo' type='Sezione'>
  <div2 xml:id='UNC11' n='1' type='Capitolo'>
    <head>I. Mia moglie e il mio naso</head>
    <p>&mdash; Che fai? &mdash; mia moglie mi domandò,
vedendomi...</p>
```

(da *Uno, nessuno e centomila* di Pirandello)

Prosa, versi, testi drammatici

PROSA

I paragrafi che costituiscono una divisione testuale dovrebbero essere marcati con `<p>`

Esempio:

```
<body>
```

```
...
```

```
<p>A seguire, non ci saremmo mai intesi; perché se a me  
stava a cuore la tigre, a lei il cacciatore. </p>
```

```
...
```

```
</body>
```

(da *Quaderni di Serafino Gubbio operatore* di Luigi Pirandello)

Versi

- `</>` contiene una singola linea, anche incompleta, di un verso. Tra gli attributi:
 - *part*: specifica se la linea è metricamente completa o meno. I valori legali sono: *F* per la parte finale di una linea incompleta, *Y* se la linea è metricamente incompleta, *N* se la linea è completa, o se nessuna assunzione è fatta sulla sua completezza, *I* per la parte iniziale di una linea incompleta, *M* per la parte mediana di una linea incompleta.
- `<lg>` contiene un gruppo di versi che costituiscono un'unità formale, per esempio una stanza, una terzina, un refrain, un paragrafo in versi, etc.

Esempio di codifica del verso

Qui, per esempio, abbiamo l'inizio di un testo poetico del quale sono marcate le linee di verso e le stanze:

```
<lg n='I'>  
  <l>Spesso il male di vivere ho incontrato:</l>  
  <l>era il rivo strozzato che gorgoglia,</l>  
  <l>era l'incartocciarsi della foglia</l>  
  <l>riarsa, era il cavallo stramazzone.</l>  
</lg>
```

(da *Ossi di seppia* di Eugenio Montale)

Si noti come l'elemento `</>` marchi il verso, non le linee tipografiche.

Testi drammatici

- `<sp>` contiene una singola battuta in un testo drammatico, o un brano presentato come tale (cioè per la recitazione) in un testo in prosa o in versi. Tra gli attributi:
 - `@who`: identifica il soggetto parlante
- `<speaker>` contiene una forma particolare di intestazione o etichetta, che fornisce i nomi di uno o più parlanti in un testo o frammento drammatico.
- `<stage>` contiene qualsiasi tipo di didascalie e direttive di scena all'interno di un testo o di un frammento di testo drammatico. Tra gli attributi:
 - `@type`: indica il tipo di direttiva di scena. I valori consigliati includono *ENTRATA*, *USCITA*, *MESSA IN SCENA*, *DIZIONE*, etc.

Esempio di *markup* dei versi in testi teatrali

```
<div1 type='Atto' n='I'><head>ATTO I</head>
  <div2 type='Scena' n='1'><head>SCENA 1</head>
    <stage rend='italic'>Francesco di scolta. Entra
    Bernardo</stage>
    <sp><speaker>Bern<l part='y'>Chi è là?
    <sp><speaker>Fran<l>Rispondi tu, prima, e manifestati.
    <sp><speaker>Bern<l part='i'>Viva il re!
    <sp><speaker>Fran<l part='m'>Bernardo?
    <sp><speaker>Bern<l part='f'>Quello.
    <sp><speaker>Fran<l>Esattamente venite alla vostr'ora.
    ...
  </div2>
..
</div1>
```

(dal primo atto, scena prima, dell'*Amleto* di W. Shakespeare)

Esempio di *markup* del dialogo in testo teatrale

```
<sp who='GUB'><speaker>Serafino Gubbio operatore</speaker>  
<p>&mdash; Che vuole che le dica? Lei in questo momento,  
  lo  
riconoscerà è molto eccitato.  
<sp who='FER'><speaker>Carlo Ferro</speaker>  
<p>&mdash; Ma posso esser calmo?  
<sp who='GUB'><speaker>Serafino Gubbio operatore</speaker>  
<p>&mdash; Ah, capisco...  
<sp who='FER'><speaker>Carlo Ferro</speaker>  
<p>&mdash; Ne ho ragione, mi sembra!  
<sp who='GUB'><speaker>Serafino Gubbio operatore</speaker>  
<p>&mdash; Sì, senza dubbio! Ma in tale stato, caro Ferro,  
è anche molto facile esagerare.</sp>
```

(da *Quaderni di Serafino Gubbio operatore* di Pirandello)

Numeri di pagina e di linea

Le interruzioni di pagina e di linea possono essere marcate con i seguenti elementi vuoti.

- `<pb/>` segnala i confini tra una pagina di un testo e la successiva in un sistema di riferimento standard.
- `<lb/>` segnala l'inizio di una nuova linea (tipografica) in una certa edizione o versione di un testo.

Questi elementi marcano un punto singolo nel testo, non una sua porzione.

L'attributo globale `@n` dovrebbe essere usato per fornire il numero della pagina o di linea che inizia con il marcatore.

Inoltre, questi due elementi condividono l'attributo `@ed` che indica l'edizione o versione in cui l'interruzione della pagina è localizzata a questo punto.

Esempio

<p>La notte era bella come nel 1914, e Parigi altrettanto minacciata. Il chiaro di luna pareva quasi un vago, ininterrotto magnesio che permettesse di fissare un'ultima volta le immagini notturne di certe belle vedute, come place

<pb ed='ED1' n='128' />

Vendôme o place de la Concorde, alle quali il mio terrore per le bombe che forse le avrebbero fra poco distrutte,

<pb ed='ED2' n='97' /> dava per contrasto, nella loro bellezza ancora intatta, una sorta di pienezza[...]</p>

(da *Il tempo ritrovato* di Marcel Proust)

Elementi 'pietra miliare'

Gli elementi `<pb/>` e `<lb/>` sono casi particolari della classe di cosiddetti elementi 'pietra miliare', che marcano i punti di riferimento all'interno di un testo.

La *TEI* prevede anche un elemento generico `<milestone>` che non è ristretto a casi particolari ma può segnalare ogni tipo di punto di riferimento: per esempio, un'interruzione di colonna, l'inizio di un nuovo tipo di sezione non marcata altrimenti, etc.

Soluzione all'overlapping?

Elementi presenti in tutti i documenti

Logic and semantic (1)

- Highlighted words or phrases in a text are commonly distinguished from their surroundings for a reason.
- Only a thorough understanding of the text and the language can lead to a correct identification and interpretation.
- The underlying semantics may be encoded with far more specific elements than the generic `<hi rend="something">`.
- Highlighting is commonly used to render the following logical and semantic features:
 - Emphasis `<emph>`, foreign words `<foreign>` and other linguistically distinct uses `<distinct>` of highlighting
 - The use of quotation marks in the representation of speech and thought `<said>`, quotation `<quote>`, cited quotation `<cit>`.
 - Technical terms `<term>`, glosses `<gloss>`

<http://tbe.kantl.be/TBE/modules/TBED01v00.htm#logicalfeatures>

Elementi presenti in tutti i documenti

Logic and semantic (2)

- **Proper nouns name, people, places, and objects** and are easily traceable in a text.
- This may be encoded with `<name>` carrying a `@type` attribute specifying the kind of object referred to.
 - `@key` - provides an externally-defined means of identifying the entity (or entities) being named, using a coded value of some kind.
 - `@ref` - (reference) provides an explicit means of locating a full definition for the entity being named by means of one or more URIs.
- However, people, places, and objects may also be referred to with common nouns, for which the element `<rs>` (for referring string) may be used. This element may also carry a `@type` attribute specifying the kind of object referred to.
- Any expression defining a date or time may be encoded with the corresponding elements `<date>` and `<time>`.
- Numbers and measures may be encoded using `<num>` and `<measure>` respectively.
- Postal and electronic addresses may be encoded by using `<address>` and `<email>` respectively.

Un esempio

JACOPO ACCIAIUOLI A VESPASIANO - Dall'accampamento aragonese, 21 luglio 1463
Firenze, Bibl. Medic. Laur., Plut. 90 sup. 30, ff. 28v-29v

Raccomandami a

`<name type="persona" ref="#Po1" key="Piero Acciaiuoli">Piero</name> et a
<name type="persona" ref="#Do1" key="Donato Acciaiuoli">Donato</name> et alla
brighata di casa, et a messer`

`<name ref="#GA1">Giovanni Argiropolo</name> et a tutta l'<name
type="organizzazione">Accademia</name>.`

Datum in `<name type="luogo" ref="#RCTG1">Regijs Castris apud Turres
Garigliani</name> <date>die 21 julij 1463</date>.`

FERDINANDO D'ARAGONA A VESPASIANO - Pozzuoli, 15 gennaio 1468
Ed. Trinchera, p. 403; Cagni, pp. 155-156

`<persname ref="#ADC1">Alfonso duca di Calabria</persname>`

[...] Quanto alla parte dello Illustrissimo

`<rs type="person" ref="#ADC1">nostro Primogenito</rs>, havemo gran piacere
habia satisfacto alli animi de quisti excelsi Signori et homini de questa`

`<rs type="place" ref="#Fil">cità</rs>, et molto li restamo obligati dello
honore che per quelli lì hè stato facto allo prefato <rs type="person"
ref="#ADC1">Duca</rs>`

Elementi presenti in tutti i documenti

Da Logic/semantic ad Analytical Features

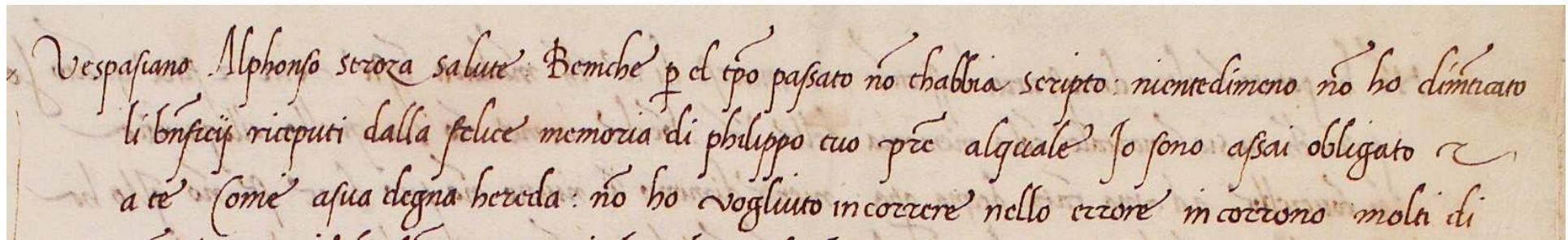
- ❑ It is sometimes useful to encode abbreviations and their expansions in texts.
- ❑ This facilitates special processing, regularisation by the full form of an abbreviation, or the rendering of different possible expansions of an abbreviation.
- ❑ Abbreviations may be marked using `<abbr>`. The `@type` attribute may be used to distinguish types of abbreviations by their function.
- ❑ Alternatively, and depending on the encoder's preference, the expansion of an abbreviation may be encoded with `<expan>`. This is often done when the editor or encoder of a text has silently expanded the abbreviation for whatever reason.
- ❑ This will commonly be combined with the `<abbr>` element inside a `<choice>` element to record the relationship between the abbreviation and its expansion

<http://tbe.kantl.be/TBE/modules/TBED01v00.htm#logicalfeatures>

Un esempio

VESPASIANO AD ALFONSO STROZZI (Antella, 8 ottobre 1493)
Firenze, Arch. di Stato, *Carte Stroziane*, serie III, vol. 145, ff. 86-87v

Vespasiano Alphonso Stroza salute. Bemché **<abbr expans="per" type="troncamento">&p-asta-tagliata;</abbr>** el t**<abbr expans="em" type="contrazione">&tratto-sopra;</abbr>**po passato no**<abbr expans="n">&tratto-sopra;</abbr>** t'habbia scripto, nientedi-meno no**<abbr expans="n">&tratto-sopra;</abbr>** ho dim**<abbr expans="en">&tratto-sopra;</abbr>**ticato li b**<abbr expans="e">&tratto-sopra;</abbr>**n**<abbr expans="e">&tratto-sopra;</abbr>**ficij riceputi dalla felice memoria di Philippo tuo **<abbr expans="padre">&pre-tratto-sopra;</abbr>**, al quale jo sono assai obligato, **<abbr expans="et" type="nota tironiana">&7;</abbr>** a te come a sua degna hereda.



Vespasiano Alphonso stroza salute. Bemche p el tpo passato no chabbia scripto: nientedimeno no ho dimtato
li bnficij riceputi dalla felice memoria di philippo tuo pre alquale? Jo sono assai obligato
a te come asua degna hereda: no ho voluto incorrere nello errore incorrono molti di

Elementi presenti in tutti i documenti

Analytical Features: *Notes and Annotations*

The most explicit form of textual annotation is the addition of notes to the text using `<note>`.

All notes should be marked using the same tag `<note>`, whether they are already present in the text or supplied by the editor, whether they appear as block notes in the main text area, at the foot of the page, at the end of the chapter or volume, in the margin, or in some other place.

The `@type` attribute distinguishes the different types of annotations in use in a text.

In a `@resp` attribute, the responsible subject for a note can be documented.

Where possible, a note can be inserted in the text at the point at which its identifier or mark first appears.

The location of the note may be documented using a `@place` attribute.

<http://tbe.kantl.be/TBE/modules/TBED01v00.htm#analyticalfeatures>

to et difficulta introducto certo ragiona-
 mento da cordo uenendo il consolo a collo-
 quio collui/ fu preso ad inganno et insieme
 cōli sedici nauī ne summato ad Annibale parte
 ua la fortuna auer fauorigiato grandemente
 i cartaginesi in questo principio. Ma po che di
 poi q̄i come uolesse far uendetta et afflixerli
 cartaginesi cō simile danno et giactua po
 che Annibale montato i speranza pla p̄sura
 del consolo et delle nauī nō prendendo exe-
 mplo degli errori altrui ma procedendo
 cō simile temerita sentendo che il resto delle
 nauī romane douea fare il cammino per loli-
 to d'italia usò la ciucilia. Lui cō cinquanta
 nauī elette et bene i punto sife i contra Et
 nauigando pe contrario cammino usò la ue-
 nimento di romani et sprouidutanti allo spu-
 ritate di certo promuntorio se scontro nel na-
 uilio de romanj dal q̄le subito torniato et
 uinto p̄de q̄i tolte le nauī cinquanta che
 seco auēua. Et la persona sua ancora fu p̄sa.
 Ma poi trasogandosi per la battaglia fuor-
 dogni speranza de nemici et de suoi sc̄apo

*Cornelio Consolo
 preso con sedici
 nauē nel porto
 de lipari dal pre-
 fetto de cartagi-
 nesi.*

*cinquantanaue
 de cartaginesi
 presi da romani*

```
<note
  target="#P1"
  place="margin-right"
  type="gloss"
  resp="#AM">

<lb/>Cornelio Consolo
<lb/> preso con <reg type="SC">
  </reg>se<choice><orig>&dTonda;
  </orig><reg
  type="SG">d</reg></choice>ici
<lb/>naue nel<reg type="SC"> </reg>porto
<lb/><choice><orig>&dTonda; </orig><reg
  type="SG">d</reg></choice>e
  <choice><orig>l</orig> <reg
  type="aggiuntaMaiuscola">L
  </reg></choice>ipari
  <choice><orig>&dTonda; </orig> <reg
  type="SG">d</reg></choice>al <reg
  type="SC"> </reg>pre
<lb/>fetto <choice><orig>&dTonda; </orig>
  <reg type="SG">d</reg></choice>e
  <reg
  type="aggiuntaPunteggiatura">' </reg>
  <reg type="SC">
  </reg><choice><orig>c</orig> <reg
  type="aggiuntaMaiuscola">C</reg>
  </choice>artagi
<lb/>nesi<reg
  type="aggiuntaPunteggiatura">.</reg>
</note>
```

Elementi presenti in tutti i documenti

Analytical Features: *Apparent Errors*

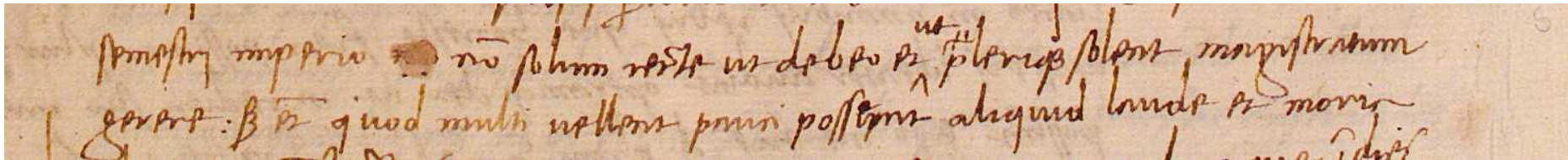
- Apparent errors in the text may be indicated using `<sic>` or corrected inside `<corr>`.
- Alternatively, the encoder may both record the original source text and provide a correction by using both `<sic>` and `<corr>` in either order wrapped in a `<choice>`.
- The encoder may encode the degree of certainty associated with the intervention or interpretation using a `@cert` attribute and indicate the agency responsible for the intervention or interpretation, for instance an editor or transcriber, using `@resp`.
 - The value of `@resp` is a pointer to an element in the document header that is associated with a person responsible for the intervention.
- The attribute value of `@resp` points to a `<name>` element in the `<teiHeader>`, for example in the `<respStmt>`.

<http://tbe.kantl.be/TBE/modules/TBED01v00.htm#errors>

Un esempio

DONATO ACCIAIUOLI A VESPASIANO (Poppi, 10 luglio 1462)
Firenze, Bibl. Naz., Magl. VIII, 1390, f. 41v

semestri imperio non solum recte, ut debeo et ut plerique solent, magistratum
gerere, sed etiam - quod multi vellent, pauci possunt - aliquid laude et
<sic corr="memoria">moria</sic>



Dalla TEI P5 è stato inserito l'attributo <choice> per documentare sia l'errore che la lezione alternativa proposta:

Es.

<choice>

<sic>moria</sic>

**<corr type="aplografia" resp="#FT" cert="100%" xml:id="015">
memoria</corr>**

</choice>

Elementi presenti in tutti i documenti

Analytical Features: *Regularisation and Normalisation*

- Standard or regularised forms for variant forms or non-standard spelling may be provided for a number of reasons.
- This is called regularisation or normalisation.
- The original, non-normalized form may be flagged using `<orig>`.
- If the encoder wants to indicate that certain words have been normalised, which means modernisation of spelling in this example, `<reg>` may be used.
- Alternatively the encoder may decide to record both the original form `<orig>` and the regularised form `<reg>` wrapped inside a `<choice>`. In the case of the modernisation of spelling, an electronic text could thus serve as the basis of an old- or new-spelling edition.
- The `@resp` attribute may be used to specify the agency responsible for the regularisation or normalisation.

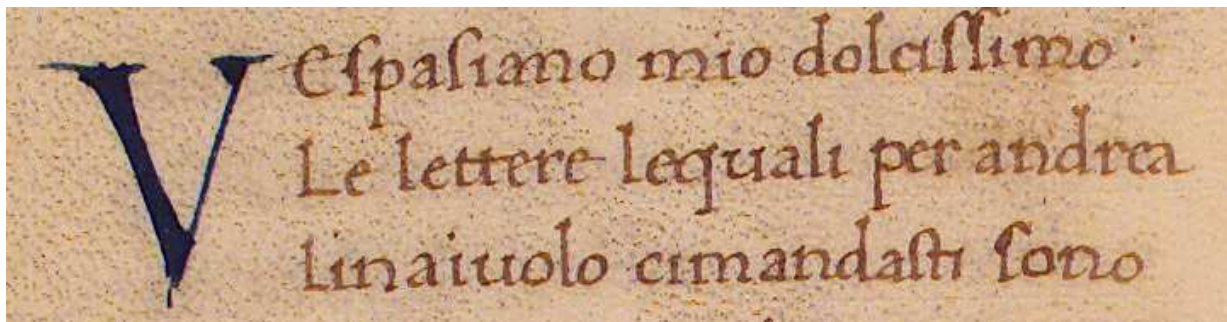
<http://tbe.kantl.be/TBE/modules/TBED01v00.htm#regularisation>

Esempio

DONATO ACCIAIUOLI A VESPASIANO

[Monte Gufone], 28 settembre 1446

Firenze, Bibl. Medic. Laur., Plut. 90 sup. 30, ff. 17v-18.



```
<hi rend="iniziale-inchiostrata">V</hi>e&slong;pa&slong;iano mio dolci&slong; &slong;imo  
<choice>  
<orig type="interpunzione-originale">:</orig>  
<reg type="sostituzione-interpunzione">,</reg>  
</choice>  
<choice>  
<orig type="maiuscola">L</orig>  
<reg type="normalizzazione">l</reg>  
</choice>e lettere  
le<choice>  
<orig type="scriptio-continua"></orig>  
<reg type="spaziatura">&spazio;</reg>  
</choice>quali per  
<choice>  
<orig type="minuscola">a</orig>  
<reg type="normalizzazione">A</reg>  
</choice>ndrea
```


Elementi presenti in tutti i documenti

Analytical Features: *Additions, Deletions, and Omissions* (1)

- ❑ Another editorial intervention in the text may be the documentation and creation of additions, deletions and omissions.
- ❑ When transcribing a source document, `<gap>` may be used to indicate a point where material has been omitted both because the material is illegible, invisible or inaudible in the source and because the editor or transcriber has decided to omit material for editorial reasons or as part of sampling practice.
- ❑ The reason for omission may be given in a `@reason` attribute.
 - Sample values include sampling, illegible, inaudible, irrelevant, cancelled.
- ❑ Additional attributes like `@extent` and `@unit` may document the amount of characters, words, lines or any other unit omitted.
- ❑ The `<gap>` element may appear as an empty element, but may also contain a description of the material omitted using `<desc>`.
- ❑ When an editor wants to mark his or her own additions as editorial interventions in the text `<supplied>`.

<http://tbe.kantl.be/TBE/modules/TBED01v00.htm#additions>

Esempio

VESPASIANO A COSIMO DE' MEDICI [Firenze, fine 1463 o inizio 1464]
Firenze, Arch. di Stato, *Mediceo avanti il Principato*, filza XII, n. 413

finita di tutto: volendo ve la mandi, ve la manderò.

<gap reason="strappo della carta">

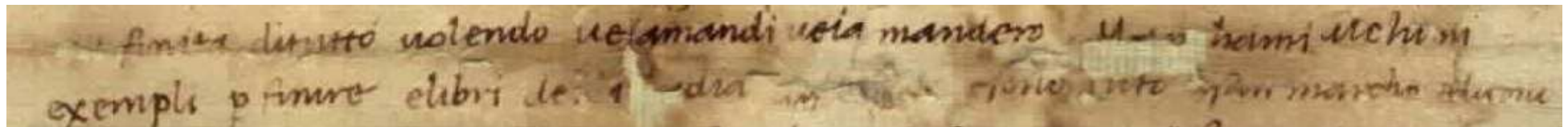
<supplied>Manc</supplied>hami

</gap> alchuni exempli per finire e libri del

<gap reason="strappo della carta e abrasione">

<supplied>la Badia di Fiesole e</supplied>

</gap>



Elementi presenti in tutti i documenti

Analytical Features: *Additions, Deletions, and Omissions* (2)

- Where words or phrases of moderate lengths have been added or deleted in the copy text.
- This may be recorded using `<add>` and ``. As with all TEI elements, information on the actual rendition of the additions and deletions can be provided in the global `@rend` attribute.
- Additionally, the place of the addition may also be recorded using `@place`.
- Where deletions in the copy text cannot be read with confidence, `<unclear>` should be used with the `@reason` attribute indicating that the difficulty of transcription is due to deletion.

<http://tbe.kantl.be/TBE/modules/TBED01v00.htm#additions>

Esempio

VESPASIANO A FILIPPO PODOCATARO – (Firenze, 4 dicembre 1448)
Firenze, Bibl. Naz., Magl. VIII, 1390, f. 102

In questo esempio la lettera, autografo di Donato Acciaiuoli, è stata corretta da altra mano, identificata in quella di Iacopo Ammannati

voluminis

~~hand="DoAc" rend="overstrike">po essent grossi sex.~~

~~rend="overstrike" hand="IaAm">litteratura~~

~~place="supralinear" hand="IaAm" type="correzione">~~

~~abbr expan="littere">&lre-tratto-superiore;</abbr> scriptoris potior~~

~~place="inline">e~~

~~place="supralinear">s tu~~

~~hand="IaAm">a~~

~~place="inline">is~~

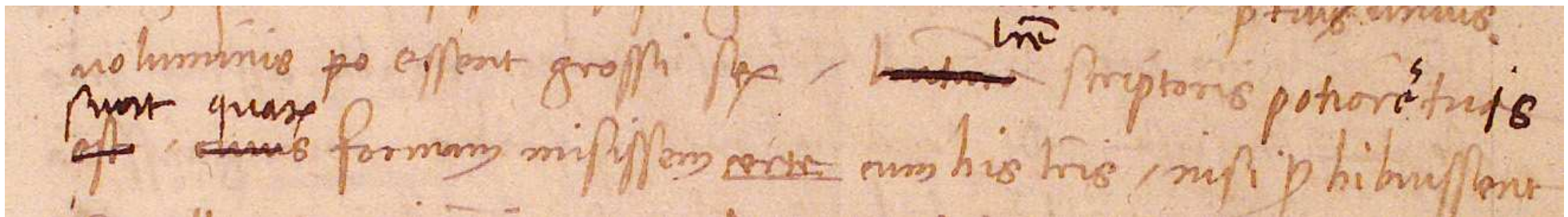
~~hand="IaAm">est~~

~~place="supralinear" hand="IaAm">sunt~~

~~hand="IaAm">cuius~~

~~place="supralinear" hand="IaAm">qua~~

~~abbr expan="rum">&cancelletto;</abbr> formam misissem recte cum his~~



Esempio

VESPASIANO A PIERO DE' MEDICI

(Firenze, 19 maggio 1458)

Firenze, Arch. di Stato, *Mediceo avanti il Principato*, filza XVII, n. 176

uno Istatio Tebaidos, di che ne resta a essere paghato, che gliene
fo richordo in sulla detta scritta

```
<damage type="macchia" agent="foxing">
```

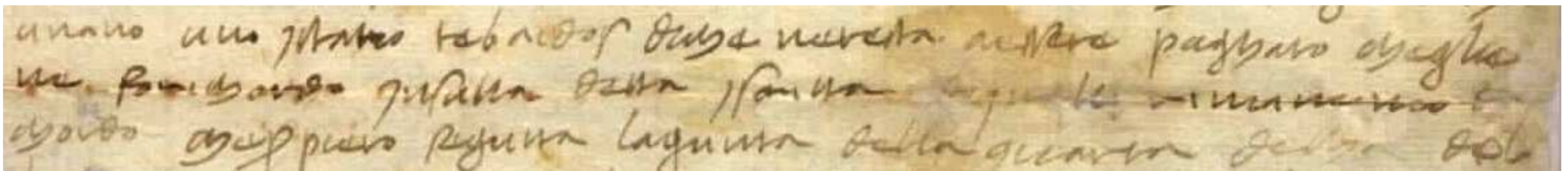
```
<unclear>
```

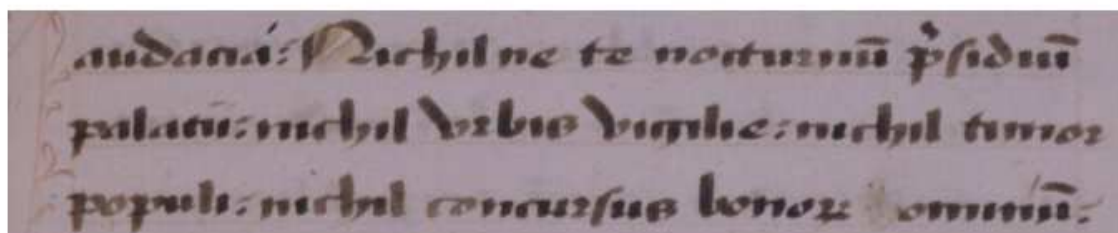
```
<supplied>di quello</supplied>
```

```
</unclear>
```

```
</damage>
```

rimanemo d'achordo. Meser Piero seguita l'aggiunta della quarta Decha





Markup XML/TEI (P4) della porzione di manoscritto in immagine

```
audacia <orig reg="?">&segno_paragrafematico;</orig>
<hi id="01">N</hi>i<sic corr="">c</sic>hilne
te nocturnu<abbr expan="m">&tratto_sopra;</abbr>
p<abbr expan="rae">&sopra_punto;</abbr>
sidiu<abbr expan="m">&tratto_sopra;</abbr>
<orig reg="P">p</orig>alatii
<orig reg=", ">&virgola_medievale;</orig>
ni<sic corr="">c</sic>hil urbi<orig reg="s">&s;</orig>
<orig reg="v">u</orig>igili<sic corr="ae">e</sic>
<orig reg=", ">&virgola_medievale;</orig>
ni<sic corr="">c</sic>hil timor
populi<orig reg=", ">&virgola_medievale;</orig>
ni<sic corr="">c</sic>hil concursu<orig reg="s">&s;</orig>
bono<abbr expan="rum">&rum;</abbr>
<del rend="rasura">h</del>
omni<abbr expan="m">&tratto_sopra;</abbr>
<orig reg=", ">&virgola_medievale;</orig>
```

LEGENDA

<orig> testo come si legge nel manoscritto;
reg dà la versione normalizzata
<sic> segnala la presenza di un errore o di un luogo su cui l'editore intende intervenire;
corr riporta la versione corretta o modificata
<abbr> indica un'abbreviazione (espressa con riferimento all'entità carattere in grado di riprodurre il segno abbreviativo);
expan scioglie l'abbreviazione
 indica una cancellazione;
rend specifica le modalità della cancellazione
<hi> segnala una qualche forma di enfasi;
id associa all'enfasi un identificativo univoco

Dal file XML possono essere prodotti due diversi output. Questo è possibile associando allo stesso documento un diverso foglio di stile. Nel caso della **trascrizione diplomatica** si rispettano i segni di scrittura, rendendoli così come compaiono nel manoscritto, tentando anche di riprodurre i fenomeni riscontrati (segni speciali); nel caso della **trascrizione critica** il testo viene sostituito con i valori degli attributi (corr per l'elemento sic, expan per l'elemento abbr, reg per l'elemento orig). Con l'ausilio di elementi tipografici (stili e colori) i fenomeni (abbreviazioni, correzioni e normalizzazioni) possono essere resi in layout.

Ipotesi di trascrizione diplomatica

audacia: Nichilne te nocturnū p̄sidiū
palatii: nichil urbiß uigilie : nichil timor
populi: nichil concursuß bonoꝝ omniū:

Esempio di trascrizione critica

audacia? Nihilne te nocturnum praesidium
Palatii, nihil urbis uigiliae, nihil timor populi,
nihil concursus bonorum omnium,

Sitografia di riferimento

- TEI P5 Guidelines
 - Intro: <http://www.tei-c.org/Guidelines/P5/>
 - Online version: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- TEI Lite (trad. it)
 - http://www.tei-c.org/Guidelines/Customization/Lite/teiu5_it.xml
- TEI by example
 - <http://tbe.kantl.be/TBE/TBE.htm>
- TEI Tutorials
 - <http://www.tei-c.org/Support/Learn/tutorials.xml>
- Biblioteca italiana
 - <http://www.bibliotecaitaliana.it/>

Grazie per l'attenzione!

francesca.tomasi@unibo.it